Enhancing the Application of Extreme Bounds Analysis in Stata

2012 Stata User Group Meeting, London

Marie Poprawe

ETH Zurich

September 2012

Motivation I

Sensitivity Analysis: Can a set of explanatory variables robustly explain a certain dependent variable?

- Leamer (1985): Sensitivity Analysis Would Help (AER) - analysis of the extreme bounds of regression coefficients
  - $$Y_i = \alpha X_i + \beta C_i + \gamma Z_i + u_i \quad (1)$$
  - Y_i is some variable of interest that one wishes to explain by the RHS variables, X_i is a vector of standard explanatory variables (all robust), C_i is an explanatory variable whose robustness needs to be tested, and Z_i is a vector of potential additional explanatory variables.

Motivation II

- **Leamer's extreme bound test** for variable C says that if the lower extreme bound of β (the lowest value of β minus two standard deviations) is negative while the upper extreme bound for β (the highest value for β plus two standard deviations) is positive, the variable C is not robustly related to Y.
- Application to growth regressions: Levine and Renelt (1992) (AER)
- In practice: Run regressions with all possible combinations of variables
- Implemented by Impavido (1998): `eba`

Motivation III

Sala-i-Martin (1997): I Just Ran Two Million Regressions (AER)

- Argues that Leamer's criterion is too strong.
- Proposal: Analyse the entire distribution of the estimates of β.
- **Sala-i-Martin's criterion:** If the average 90% confidence interval of a regression coefficient does not include zero then the variable is correlated with Y.
- Runs cross-country growth regressions to test which variables robustly explain growth
- *This has not been implemented in Stata yet*

	Number of Regressions = 30,856 NAME OF TESTED VARIABLE	(1) Lower Extreme	(2) Upper Extreme	(3) Fraction Significant	(4) Beta	(5) Standard Deviation	(6) CDF Normal	(7) CDF Non-Normal (Weighted)	(8) CDF Non-Normal (Not Weighted)
1	Equipment Investment	-0.0396	0.5268	99.97%	0.21748	0.04081	1.000	1.000	1.000
2	Number of Years Open Economy	-0.0025	0.0438	99.97%	0.01948	0.00424	1.000	1.000	1.000
3	Fraction of Confucious	0.0038	0.1266	100.00%	0.06757	0.01492	1.000	1.000	1.000
4	Rule of Law	-0.0124	0.0599	92.04%	0.01895	0.00492	1.000	1.000	0.993
5	Fraction of Muslim	-0.0155	0.0368	88.67%	0.01421	0.00353	1.000	1.000	0.988
6	Political Rights	-0.0134	0.0077	33.73%	-0.00265	0.00087	0.999	0.998	0.926
7	Latin American Dummy	-0.0354	0.0112	88.28%	-0.01154	0.00291	1.000	0.998	0.986
8	Sub-Sahara African Dummy	-0.0377	0.0174	76.38%	-0.01212	0.00322	1.000	0.997	0.978
9	Civil Liberties	-0.0113	0.0130	16.90%	-0.00290	0.00102	0.998	0.997	0.848
10	Revolutions and Coups	-0.0377	0.0358	2.81%	-0.01179	0.00452	0.995	0.995	0.696
11	Fraction of GDP in Mining	-0.1654	0.1361	17.18%	0.03533	0.01383	0.995	0.994	0.729
12	S.D. Black Market Premium	-0.0001	0.0000	35.29%	-0.00003	0.00001	0.993	0.993	0.955
13	Primary Exports in 1970	-0.0450	0.0167	93.59%	-0.01399	0.00526	0.996	0.990	0.993
14	Degree of Capitalism	-0.0036	0.0089	51.03%	0.00184	0.00079	0.990	0.987	0.944
15	War Dummy	-0.0168	0.0126	17.09%	-0.00562	0.00233	0.992	0.984	0.870
16	Non-Equipment Investment	-0.0633	0.2468	76.32%	0.05622	0.02424	0.990	0.982	0.978
17	Absolute Lattitude	-0.0004	0.0009	66.20%	0.00023	0.00009	0.993	0.980	0.965
18	Exch. Rate Distortions	-0.0003	0.0001	54.29%	-0.00006	0.00003	0.975	0.968	0.958
19	Fraction of Protestant	-0.0480	0.0172	57.07%	-0.01286	0.00525	0.993	0.966	0.958
20	Fraction of Buddhist	-0.0142	0.0554	92.47%	0.01485	0.00755	0.975	0.964	0.994
21	Fraction of Catholic	-0.0305	0.0120	84.18%	-0.00891	0.00341	0.996	0.963	0.984
22	Spanish Colony	-0.0258	0.0286	45.42%	-0.00647	0.00321	0.978	0.938	0.889
23	Public Investment Share	-0.2309	0.2714	1.16%	0.04070	0.02758	0.930	0.915	0.691
24	Frac. Pop. Spk. English	-0.0350	0.0166	27.29%	-0.00682	0.00480	0.922	0.910	0.923
25	Defense Spending Share	-0.2120	0.3269	16.30%	-0.06417	0.04693	0.914	0.909	0.883
26	Age	-0.0002	0.0001	15.10%	-0.00004	0.00003	0.907	0.903	0.918
27	Public Consumption Share	-0.2754	0.1040	35.60%	-0.02205	0.01655	0.909	0.868	0.920
28	Average Inflation Rate 60-90	-0.0010	0.0005	6.56%	-0.00002	0.00002	0.859	0.856	0.752
29	Size Labor Force (Scale Effect)	-0.0003	0.0008	0.43%	0.00005	0.00005	0.846	0.835	0.744
30	Frac. Pop. Spk. Foreign Language	-0.0162	0.0245	19.43%	0.00486	0.00304	0.945	0.831	0.843
31	Black Market Premium	-0.0228	0.0381	3.36%	-0.00391	0.00356	0.864	0.825	0.707
32	S.D. Inflation 60-90	-0.1664	0.2322	3.81%	-0.00415	0.00465	0.814	0.811	0.560
33	Growth Rate of Population	-1.2817	1.1651	0.31%	0.20409	0.18643	0.863	0.807	0.531
34	Ratio Workers to Population	-0.0491	0.0456	21.75%	0.00482	0.00530	0.819	0.766	0.773
35	Fraction of Jewish	-1.8490	1.0118	0.16%	0.00923	0.01371	0.750	0.747	0.617
36	Liquid Liabilities to GDP	-0.0335	0.0802	69.32%	0.00819	0.01012	0.791	0.735	0.962
37	Avg. Years of Primary School	-3.5131	7.9705	3.91%	-0.00068	0.00117	0.720	0.704	0.811
38	French Colony	-0.0238	0.0306	0.11%	0.00177	0.00315	0.713	0.702	0.650
39	Political Assassinations	-0.1833	0.1587	0.02%	0.01338	0.02534	0.701	0.697	0.675
40	S.D. Domestic Credit	-0.0004	0.0002	4.75%	-0.00001	0.00001	0.696	0.696	0.715
41	H*log(GDP60)	-0.0055	0.0040	2.25%	-0.00006	0.00011	0.697	0.689	0.688
42	Fraction of Hindu	-0.3100	0.1018	0.50%	0.00306	0.00764	0.656	0.654	0.524
43	Avg. Years of Schooling = H	-7.9734	3.5109	0.85%	-0.00038	0.00100	0.646	0.653	0.623
44	Secondary School Enrollment	-0.0598	0.0771	2.47%	-0.00438	0.01058	0.661	0.649	0.711
45	Ethnolinguistic Fractionalization	-0.0302	0.0253	0.64%	-0.00183	0.00457	0.655	0.643	0.614
46	Outward Orientation	-0.0097	0.0157	2.51%	-0.00081	0.00233	0.635	0.634	0.794
47	Index of Democracy 1965	-0.0402	0.0233	17.73%	-0.00176	0.00498	0.638	0.633	0.890
48	Tariff Restrictions	-0.5715	0.4344	0.53%	0.01793	0.05657	0.624	0.624	0.670
49	Free Trade Openness	-0.1042	0.1490	3.84%	-0.00568	0.01893	0.618	0.617	0.818
50	Avg. Years of Higher School	-3.5045	7.9721	0.01%	-0.00357	0.01416	0.600	0.597	0.643
51	Avg. Years of Sec. School	-3.5069	7.9789	2.94%	0.00061	0.00259	0.593	0.592	0.800
52	Political Instability	-0.0684	0.1024	0.30%	-0.00193	0.01059	0.572	0.581	0.588
53	Gov. Education Spending Share	-0.6854	0.7519	0.53%	0.02822	0.12604	0.589	0.580	0.677
54	Higher Educ. Enrollment	-0.1833	0.2323	0.01%	-0.00692	0.03290	0.583	0.579	0.543
55	British Colony	-0.0180	0.0133	1.25%	-0.00047	0.00232	0.581	0.579	0.612
56	Urbanization Rate	-0.0438	0.0520	1.01%	-0.00156	0.00825	0.575	0.577	0.746
57	Growth of Domestic Credit 60-90	-0.0005	0.0008	0.02%	-0.00001	0.00004	0.565	0.565	0.542
58	Area (Scale Effect)	-0.0031	0.0043	0.02%	0.00005	0.00064	0.532	0.532	0.539
59	Terms of Trade Growth	-0.3437	0.2348	0.05%	0.00129	0.04117	0.512	0.511	0.628
	Number of Regressions = 32509 VARIABLES NOT TESTED								
	log(GDP per capita 1960)	-0.0336	0.0009	99.98%	-0.01325	0.00230	1.000	1.000	1.000
	Primary School Enrollment, 1960	-0.0384	0.0680	47.58%	0.01793	0.00683	0.996	0.992	0.899
	Life Expectancy, 1960	-0.0008	0.0029	96.30%	0.00083	0.00023	1.000	0.999	0.996

The advantages of enhancedeba I

eba-function:

- Only Leamer's EBA
- Up to 4 explanatory variables
- Only OLS, no options (e.g. robust standard errors)

enhancedeba-function:

- Can use both Leamer's and Sala-i-Martin's methods
- Applicable to any dataset, cross-sectional or panel
- Not just combinations of four variables, but any combinations possible
- Run any kind of regression, not just OLS

The advantages of enhancedeba II

Avoid running regressions multiple times

- Usually:
  - To test x_1: run regressions *reg* y x_1 x_2, *reg* y x_1 x_3, *reg* y x_1 x_4
  - To test x_2: run regressions *reg* y x_2 x_1, *reg* y x_2 x_3, *reg* y x_2 x_4
  - `enhancedeba` avoids running the same regression twice by running all regressions and later picking out the ones that contain each variable that needs testing.

Enhancedeba I

Syntax: `enhancedeba` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, `x(`*varlist*`) combinations(#) cmd options() preoptions() level(#) noupto logfile() printlog leamer vif(#) outputfile`]

Option	Explanation
x(varlist)	variables that should be included in every regression
combinations()	number of variables (n) in each combination
cmd()	type of regression to be run (e.g. `xtreg`)
options()	regression-type specific options (e.g. `fe`)
preoptions()	for commands where the options come before the comma (e.g. `xi:`)
level()	confidence interval
noupto	combinations of n variables or up to n variables
logfile()	log file to record EBA results
printlog	progress of EBA appears in the results window
outputfile	results written to CSV
leamer	uses Leamer methodology
vif	Variance Inflation Factor used as a criterion
onlysignificant	only significant observations are taken into account

Examples I

```
. clear

. sysuse auto
(1978 Automobile Data)

. enhancedeba price headroom trunk weight, comb(3) opt(robust) x(length) level(85)
```

```
regression command:         regress price length indepvar(s)     , robust

Confidence Interval used:   85% (.15 signficance level)
Dependent variable:         price
Independent variable(s):    headroom trunk weight
x variable(s):              length

A total of 7 regression(s) were run.
Every explanatory variable is in 4 regression(s).
```

	min beta	max beta	avg beta	avg se	% sign	CDF(0)
headroom	-711.5679	-486.4429	-597.9757	358.7918	100	100
trunk	.8252001	114.0859	56.41261	81.59962	0	100
weight	4.674047	4.753066	4.711944	1.748497	100	100
_cons	-4942.297	11488.47	4158.896	4236.28	100	57.14286

Results for the x variable(s):

	min beta	max beta	avg beta	avg se	% sign	CDF(0)
length	-102.6652	67.31455	-29.50167	39.80032	100	57.14286

Examples II

```
. enhancedeba_examples xtreg
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)
       panel variable:  idcode (unbalanced)

enhancedeba ln_wage wks_ue union tenure hours, cmd(xtreg) opt(fe) comb(2)
```

```
regression command:         xtreg ln_wage  indepvar(s)     , fe

Confidence Interval used:   95% (.05 signficance level)
Dependent variable:         ln_wage
Independent variable(s):    hours tenure union wks_ue

A total of 10 regression(s) were run.
Every explanatory variable is in 4 regression(s).
```

	min beta	max beta	avg beta	avg se	% sign	CDF(0)
hours	-.0016347	.0006786	-.0002216	.0002705	50	50
tenure	.0262867	.0357629	.032627	.000725	100	100
union	.098657	.1400584	.1154515	.0078037	100	100
wks_ue	-.0018697	-.0006076	-.0014642	.0003725	75	100
_cons	1.541016	1.788804	1.642975	.0058418	100	100

Examples III

```
. sysuse auto
(1978 Automobile Data)

. enhancedeba price mpg rep78 headroom trunk weight length, x(foreign) comb(3) leamer vif(5)
```

regression command: regress price foreign indepvar(s) ,

Confidence Interval used: 95% (.05 signficance level)
Dependent variable: price
Independent variable(s): headroom length mpg rep78 trunk weight
x variable(s): foreign

A total of 41 regression(s) were run.
Every explanatory variable is in 16 regression(s).

	min beta	max beta	avg beta	avg se	% sign	CDF(0)
headroom	-645.1354	491.5752	-396.7793	410.2221	0	87.5
length	-99.34391	100.3555	28.78662	25.41487	100	68.75
mpg	-303.8203	27.32371	-154.2233	71.68815	43.75	75
rep78	-67.84385	458.6877	164.3009	380.9681	0	87.5
trunk	-88.30739	328.4497	71.85636	93.0781	25	50
weight	3.320737	6.096515	4.29752	.6694643	100	100
_cons	-12117.64	12684.65	-620.8968	2897.106	60.97561	53.65854

Results for the x variable(s):

	min beta	max beta	avg beta	avg se	% sign	CDF(0)
foreign	-205.6112	3711.123	2451.23	788.4066	75.60976	97.56098

Examples IV

Leamer Results:

	beta	p-value	C-I low	C-I high	VIF	Obs
headroom_min	-645.1354	.9734857	-1298.923	8.652284	1.30516	74
headroom_max	491.5752	.1274705	-362.2356	1345.386	1.094511	74
length_min	58.79209	.0122506	7.776386	109.8078	3.776113	69
length_max	100.3555	1.16e-07	65.30391	135.4071	1.845836	74
mpg_min	-303.8203	.9999987	-422.4626	-185.1781	1.33706	74
mpg_max	27.32371	.3627794	-127.2085	181.8559	3.104604	69
rep78_min	-67.84385	.5629026	-918.8291	783.1414	1.559942	69
rep78_max	458.6877	.127819	-339.1808	1256.556	1.622385	69
trunk_min	-88.30739	.8685748	-244.2787	67.66392	1.83222	74
trunk_max	328.4497	.001232	119.7532	537.1462	1.879773	74
weight_min	3.320737	1.31e-12	2.531752	4.109721	1.541895	74
weight_max	3.761171	2.76e-07	2.39604	5.126303	4.570416	74
foreign_min	-205.6112	.5845368	-2117.983	1706.761	1.540226	69
foreign_max	3711.123	3.57e-07	2348.271	5073.976	1.597047	74

Conclusion I

- New implementation of Extreme Bounds Analysis approach by Leamer(1985) and Sala-i-Martin(1997)
- very flexible with respect to the type of regression, options, number of variables in each combination etc.
- No reason to apply this methodology only to growth regressions
- Can be used for any question of the form "What explains Y?" or "Can X truly explain Y?

Problems I

- Have to save results to file - this is what takes a long time
- Cannot write to macros
- Sometimes "too many literals"
- Ideas?